AMENDMENTS TO THE CLAIMS

Please amend claims 1, 16, 19, 22, and 23 as follows:

1. (Currently Amended) A method for scheduling a resource to service a plurality of pending requests received from a plurality of schedulable entities, ~~while preventing each schedulable entity from exceeding a maximum quality of service allocated to each schedulable entity,~~ comprising:

selecting a request associated with a schedulable entity, the schedulable entity being associated with a maximum allocation of the resource, the maximum allocation being specified as a maximum quality of service;

responsive to determining that servicing the selected request will exceed the schedulable entity's maximum quality of service, advancing a virtual time for scheduling the requests, without servicing the request; and

responsive to determining that servicing the selected request ~~does~~ will not exceed the schedulable entity's maximum quality of service, servicing the request and advancing the virtual time.

2. (Original) The method of claim 1, wherein the request includes a request to allocate disk space.

3. (Original) The method of claim 1, wherein the request includes a request to allocate memory.

4. (Original) The method of claim 1, wherein the request includes a request for network bandwidth.

5. (Original) The method of claim 1, wherein the request includes a request for CPU processing cycles.

6. (Original) The method of claim 1, wherein the request is selected using a fair-share scheduling algorithm.

7. (Original) The method of claim 6, wherein the fair-share scheduling algorithm is a weighted fair-share scheduling algorithm, each weight corresponding to a schedulable entity's minimum quality of service allocation.

8. (Original) The method of claim 7, wherein the minimum quality of service allocated to each schedulable entity is a minimum percentage share of the resource.

9. (Original) The method of claim 6, wherein the fair-share scheduling algorithm is a hierarchical fair-share scheduling algorithm.

10. (Original) The method of claim 6, wherein the fair-share scheduling algorithm is a hierarchical weighted fair-share scheduling algorithm, each weight corresponding to a schedulable entity's minimum quality of service allocation.

11. (Original) The method of claim 6, wherein the fair-share scheduling algorithm is a start-time fair queuing algorithm with virtual time scheduling.

12. (Original) The method of claim 1, wherein each request includes a requested duration, the method further including:

limiting the requested duration of the request to a pre-determined request duration upper bound.

13. (Original) The method of claim 1, wherein the maximum quality of service allocated to each schedulable entity is a maximum percentage share of the resource.

14. (Original) The method of claim 1, wherein a rate controller determines if servicing the request will exceed the schedulable entity's maximum quality of service.

15. (Original) The method of claim 14, wherein if the rate controller determines that servicing the request will exceed the schedulable entity's maximum quality of service, the request remains pending.

16. (Currently Amended) A method for scheduling resource requests from a plurality of schedulable entities, wherein each resource request includes a requested duration and each schedulable entity has a maximum resource allocation, the maximum resource allocation being specified as a maximum quality of service guarantee, the method comprising:

assigning a start number tag to a resource request using a start-time fair queuing algorithm with virtual time scheduling;

selecting the resource request with the smallest start number tag, the selected request having an associated schedulable entity;

limiting the requested duration of the selected resource request to a pre-determined duration upper bound;

servicing the selected resource request if servicing the selected resource request will not exceed the associated schedulable entity's maximum quality of service guarantee; and

advancing a virtual time value.

17. (Original) The method of claim 16, further including:

updating the start number tag for a resource request associated with the schedulable entity that made the selected resource request if the selected resource request is not serviced.

18. (Original) The method of claim 16, further including:

leaving the selected resource request pending if servicing the selected resource request will exceed the schedulable entity's maximum quality of service guarantee.

19. (Currently Amended) A system for scheduling pending resource requests from a plurality of schedulable entities while limiting a maximum resource allocation, the maximum resource allocation being specified as a maximum quality of service allocated to each schedulable entity, comprising:

a plurality of schedulable entity queues for holding pending resource requests, each schedulable entity queue holding resource requests from a schedulable entity;

a scheduler for selecting resource requests from the plurality of schedulable entity queues using a fair-share scheduling algorithm, and further adapted to increment a virtual time value each time a resource request is selected; and

a plurality of rate controllers associated with the plurality of schedulable entity queues, each rate controller adapted to limit the rate at which resource requests selected by the scheduler are serviced to the schedulable entity's maximum quality of service.

20. (Original) The system of claim 19, wherein each rate controller is further adapted to:

monitor the servicing of resource requests from the rate controller's associated schedulable entity queue to calculate the quality of service received by the schedulable entity; and

block the servicing of a selected resource request if the schedulable entity's maximum quality of service would be exceeded if the selected resource request was serviced.

21. (Original) The system of claim 19, wherein each schedulable entity queue is associated with a weight, and the scheduler uses a weighted fair-share queuing algorithm.

22. (Currently Amended) A hierarchical system for scheduling resource requests from a plurality of child schedulable entities while limiting ~~the~~ a maximum ~~quality of service~~ resource allocation allocated to a plurality of parent schedulable entities, the maximum resource allocation being specified as a maximum quality of service, comprising:

a plurality of child schedulable entity queues for holding pending resource requests, each child schedulable entity queue holding resource requests from a child schedulable entity;

one or more child schedulers for selecting resource requests from the plurality of child schedulable entity queues using a fair-share scheduling algorithm, and further adapted to transmit selected resource requests to a parent schedulable entity queue;

a plurality of parent schedulable entity queues, each parent schedulable entity queue receiving resource requests from a subset of the child schedulable entity queues, each parent schedulable entity queue holding resource requests received from one of the child schedulers;

a parent scheduler for selecting resource requests from the plurality of parent schedulable entity queues using a fair-share scheduling algorithm, and further adapted to increment a virtual time value each time a resource request is selected; and

a plurality of rate controllers associated with the plurality of parent schedulable entity queues, each rate controller adapted to limit the rate at which resource requests selected by the parent scheduler are serviced to a parent schedulable entity's maximum quality of service.

23. (Currently Amended) A computer program product for scheduling a plurality of pending requests for service from a resource received from a plurality of schedulable entities, while preventing each schedulable entity from ~~exceeding~~ receiving an amount of the resource that exceeds a maximum ~~quality of service~~ resource allocation allocated to each schedulable entity, the maximum resource allocation being specified as a maximum quality of service, the computer program product comprising: a computer readable medium that stores program code, including:

program code that selects a request associated with a schedulable entity using a fair-share scheduling algorithm;

program code that services the request if a rate controller determines that servicing the request will not exceed the associated schedulable entity's maximum quality of service; and

program code that advances a virtual time in the fair-share scheduling algorithm.

24. (Original) The computer program product of claim 23, wherein the fair-share scheduling algorithm is a weighted fair-share scheduling algorithm, each weight corresponding to a schedulable entity's minimum quality of service allocation.

25. (Original) The computer program product of claim 23, wherein each request includes a requested duration, the computer program product further including:

program code that limits the requested duration of the request to a pre-determined request duration upper bound.